Punishment, inequality, and welfare: a public good experiment

David Masclet, Marie Claire Villeval

Punishment, Inequality, and Welfare: A Public Good Experiment

David Masclet and Marie Claire Villeval

Abstract

This paper reports the results of an experiment that investigates the two-sided relationship between punishment and welfare. First, it contributes to the literature on the behavioral determinants of punishment by examining the role of relative income and income comparisons as a determinant of punishment in a two-stage public good game when inequality arises endogenously from the subjects' behavior. Second, this paper investigates the impact of punishment on both absolute and relative incomes. We compare three treatments of our game. The Unequal Cost treatment replicates Fehr and Gächter (2000)'s experiment under a stranger matching protocol. The Equal Cost treatment is identical to the previous one except that the ratio between the cost of one punishment point to the punisher and its cost to the target equals one. The third treatment is similar to the second one except that a partner matching protocol is implemented in order to isolate strategic motives for punishment. Our results indicate that subjects punish *even* when they cannot alter the current distribution of payoffs. We also find that in all treatments, the intensity of punishment increases in the level of inter-individual inequality. Finally, despite its cost, punishment progressively improves welfare in association with a decrease in the aggregate level of inequality over time.

JEL-Codes: A13, C92, D63

Keywords: Public Good Experiment, Punishment, Inequality Aversion, Free-Riding, Welfare.

David Masclet: CNRS, CREM, 7, place Hoche 35065 Rennes, France. Tel : +33 223 23 33 18. Email : david.masclet@univ-rennes1.fr

Marie-Claire Villeval: CNRS, GATE, 93, chemin des Mouilles 69130 Ecully, France, and Institute for the Study of Labor (IZA), Bonn, Germany. Tel: +33 472 86 60 79. Email : villeval@gate.cnrs.fr

Acknowledgements: This paper has benefited from useful comments from participants at the 2005 World Conference of the Society of Labor Economists and European Association of Labour Economists in San Francisco, the ESA Meeting in Montreal, and the 2nd International Meeting on Experimental and Behavioral Economics in Valencia, in particular from E. Fatas and L. Putterman. We thank L. Levy-Garboua for his remarks. We also thank two anonymous referees for their helpful comments. We are also grateful to R. Zeiliger for programming the experiment presented in this paper. This research has been supported by a grant from the MiRe – DREES (French Ministry of Social Affairs).

1. INTRODUCTION

Social comparisons and their consequences on decision behavior and satisfaction have inspired a huge literature (Clark and Oswald 1996; Neumark and Postlewaite 1998; Brown, Gardner, Oswald and Qian 2005; Ferrer-i-Carbonell 2005). In particular, the observation of disadvantageous inequality has been shown to generate negative reactions. We can find several examples across many social science disciplines of the consequences of disadvantageous inequality on inter-group conflicts, including tax revolt (Lowery and Siegelman, 1981), riots, civil unrest or even revolutions (Wright, Taylor and Moghaddam, 1990; Besançon, 2005; MacCulloch and Pezzini, 2004, 2007). Inequality may also trigger a willingness to hurt through sanctions and punishment, especially in within-group situations such as work teams or partnerships. However, the relationship between inequality and the individual decision to sanction is often evoked as a hypothesis but little studied in the literature.

Several experiments inspired by Fehr and Gächter's (2000) have shown that punishment is a powerful device for promoting cooperation among individuals in a public good game (Anderson and Putterman (2003), Carpenter (2007), Carpenter, Matthews and Ong'ong'a (2004), Egas and Riedl, (2005), Fehr and Gächter (2002), Masclet, Noussair, Tucker and Villeval (2003), Nikiforakis and Normann (2005), Nikiforakis, Normann and Wallace (2005), Bochet, Page, and Putterman (2007)). Sanctions appear as an effective means of alleviating the free-rider problem. In contrast, Fehr and Rockenbach (2003) identify a detrimental effect of sanctions on cooperation and a crowding-out of norm-based motivations by punishment is also observed by Houser, Xiao, McCabe, and Smith (2006). Therefore, all these studies find that subjects are

willing to engage in costly punishment but there is no consensus on the overall effect of sanctions on welfare and payoff distribution.

Our paper contributes to the literature on altruistic punishment by examining the two-sided relationship between punishment and welfare in the context of a public good experiment. First, we investigate to what extent relative income explains punishment when inequality arises endogenously from the subject's behavior. Second, we study how sanctions affect welfare and payoff distribution over time.

The first aim of the paper is to contribute to the literature on the behavioral determinants of punishment by examining the implication of income comparisons as an explanation of punishment in the framework of a two-stage public good experiment. Indeed, the role of income comparisons and inequality among the determinants of costly punishment remains unclear, in particular in the absence of strategic motivations. Two non-strategic motives are generally evoked in the literature to explain why subjects may be willing to punish. A first non-strategic motive is related to negative emotions, such as anger and disapproval. It relies on the idea that people react to unfair intentions by sacrificing a part of their payoffs in order to punish others, even when there are no reputation gains from doing so (Rabin (1993), Falk and Fischbacher (2006)). A bunch of recent papers use various methods to measure the intensity of emotions associated with the decision to punish (see for example Quervain et al., 2004; Hopfensitz and Reuben, 2005). A second non-strategic reason to punish group members relies on distributional concerns such as inequality aversion (Fehr and Schmidt (1999), Falk, Fehr and Fischbacher (2005), Bolton and Ockenfels (1999)).[1]

In this paper, we conjecture that income comparisons may affect the decision to punish in two ways. First, consistent with equity models, subjects may be willing to sanction in order to reduce income differences. In this case, individuals with distributional concerns who suffer from disadvantageous inequality would be willing to pay to punish defectors in order to reduce earnings inequality, *only* if the cost they bear is smaller than the impact of sanctions on the target's payoff. Second, it might also be the case that, even if sanctions cannot change the distribution of earnings, the individuals' decision to punish may be driven by interpersonal income comparisons. Indeed, individuals may be willing to punish those whose decisions generated unfair payoff differences because such interpersonal comparisons raise emotions that trigger sanctions. In this case, one should observe that individuals also punish, even when the cost of one punishment point to the punisher is equal to the cost of this point to the target.

The second aim of this paper is to study the opposite side of the relationship between punishment and income by investigating the impact of sanctions on the evolution of both absolute and relative incomes over time. Indeed, even if sanctions are not motivated by a willingness to reduce current payoff differences, they can affect both welfare and payoff distribution. In particular, we investigate two potential opposite effects of punishment on both welfare and inequality. Sanctions may improve welfare and reduce inequality by inciting free riders to contribute more over time. However this effect may be counterbalanced by the fact that punishment destroys resources by imposing a direct cost on both the target and the punisher, which may also increase payoff dispersion between those who incurred those costs and the others in the group.

Our experiment consists of three treatments of a two-stage public good game. In the first stage, subjects contribute voluntarily to the funding of a public good; in the second stage, after being

informed on the contribution of each group member, they may attribute costly punishment points. The three treatments only differ in the second stage of the game. Our first treatment, called Unequal Cost treatment, replicates the experiment of Fehr and Gächter (2000) under a stranger matching protocol.[2] In this treatment, punishment can directly affect payoff inequality between the punisher and the target since the ratio between the cost of one punishment point to the punisher and the cost of this point to the target is smaller than one.[3] In the second treatment, called Equal Cost treatment, the ratio between the cost of one punishment point to the punisher and the cost of this point to the target is fixed and equals one. Therefore, punishment cannot affect the current inequality level between the punisher and the target. The third treatment is equivalent to the Equal Cost treatment except that we use a partner matching protocol instead of a stranger matching protocol in order to isolate strategic motives for punishment.

Our paper is related to Falk, Fehr and Fischbacher (2005) that analyzes the importance of fairness and spite in punishment in several games, including a three-player one-shot Prisoner's Dilemma game.[4] In their experiment, spiteful punishment is carried out by subjects who value the target's payoff negatively, irrespective of the distribution of pre-punishment payoffs. The authors conclude that punishment is not primarily driven by the willingness to change payoff shares, but by the desire to retaliate and harm those who behave unfairly. In a public good environment, other papers compare the effectiveness of sanctions by varying the ratio between the cost of punishment to the punisher and its cost to the target. In particular, Anderson and Putterman (2006) show that the decision to punish obeys the law of demand; a secondary result is that individuals punish even when punishment is more costly to them than to the punished group members. Egas and Riedl (2005) vary both the cost and the impact of punishment. They observe

that cooperation is sustained only in case punishment is cheap and has a high impact on the target. Finally, the authors confirm that the decision to punish is based on both cost-benefit analysis and emotional reactions. Nikiforakis and Normann (2005) find that the contributions to the public good increase monotonically in the effectiveness of punishment, captured by the factor by which the punishers reduce the punished player's earnings; punishment increases welfare only when the factor is higher than two. Finally, Nikiforakis, Normann and Wallace (2005) investigate the role of the asymmetry in the distribution of punishing abilities within a group and they find no effect of such an asymmetry in the power to punish.

Our investigation allows us to consider several issues that earlier studies leave unaddressed. In particular the main contribution of this paper is to add more details for understanding the relationship between welfare and punishment. A first issue is whether relative income and inter-individual income comparisons matter in the decision to punish even if sanctioning cannot influence inequalities. Indeed most of the previous studies conclude that theories of inequality aversion cannot explain why subjects punish when the ratio of punishment (i.e the ratio between the cost of punishing and the cost of receiving punishment) equals one. In this paper we qualify these conclusions by stating that incomes comparisons can be at the origin of punishment even if punishment cannot reduce inequalities directly. A second issue examines the impact of punishment on welfare over time. If sanctions cannot affect the current distribution of earnings, they may have an impact on welfare and on the reduction of inequality over time, notably by inciting free riders to increase their future contributions. We therefore provide a dynamic analysis of the relationship between punishment, inequality and welfare.

We first find that punishment is not mainly driven by the willingness to reduce the current level of inequality between the punisher and the target. Indeed, individuals punish *even* in the Equal Cost treatment. Second, our results support an indirect effect of inter-individual comparisons on punishment since in all treatments, the intensity of sanctions is strongly correlated with the first-stage contribution and earning differences between the punisher and the target. This result is consistent with the fact that income differences may trigger emotions that would raise punishment even when individuals cannot reduce inequality. Last, punishment reduces inequality over time, by inciting free riders to increase their effort in future periods. Decreasing inequality is also observed when punishment is not possible but this is associated with a decrease in earnings; in contrast, when punishment is available, the reduction in the level of inequality goes along with an increase in welfare.

The remainder of the paper is organized as follows. Section 2 details our experimental design and presents the theoretical predictions of the model, with either purely selfish agents or in the presence of agents with distributional concerns. The results of the study are presented in section 3, and section 4 gives our concluding remarks.

2. EXPERIMENTAL DESIGN AND THEORETICAL PREDICTIONS

Design. The experiment is based on a public good game, involving groups of four subjects. It consists of 30 periods, divided into three segments. In the first ten periods and the last ten periods, subjects play a standard public good game. This no-punishment condition serves as a benchmark for the condition with punishment opportunities that occurs from periods 11 to 20. At

the beginning of each period, each group member is endowed with 20 units. Each member simultaneously selects a fraction of her endowment to contribute to a group account, while keeping the remainder in her private account. All funds in the group account pay a positive return to each member. The parameters are chosen so that full free riding is a dominant strategy whereas full contribution to the public good corresponds to the social optimum.

Consider first the condition without punishment. In periods 1-10 and 21-30, each subject i chooses a fraction g_i of her endowment as a contribution to the public good. All group members' decisions regarding g_i are made simultaneously. The marginal per capita return from a contribution to the group account is 0.4. Subject i 's payoff is given by:

$$\pi_i = 20 - g_i + .4\sum_{j=1}^{4} g_j \qquad (1)$$

The group members are informed of both the amount of the group contribution and their individual payoff.

Now consider the condition with punishment. Each period 11-20 of both treatments consists of a two-stage game. The first stage is identical to that in the previous condition: each group member receives an endowment and has to decide how much to invest in the group account. In the second stage, each subject, after being informed about each other group member's contribution, can assign 0 to 10 punishment points to any of the other three group members. Assigning points is costly. The main difference between the Unequal Cost treatment conducted under a stranger matching protocol (UCS) and the Equal Cost treatment also conducted under the same matching protocol (ECS) lies precisely in the monetary consequences of punishment points. The schedule of costs in these two treatments is given in Table 1.

[Table 1 about here]

In the Unequal Cost treatment (UCS), each punishment point received from any other subject reduces first-stage earnings by 10%, up to a maximum of 100%. This treatment intends to replicate that in Fehr and Gächter (2000) in order to be able to compare behavior in the base treatment. As usual, contributions are listed in random order and with a different identification number on the screen each period so that it is impossible to target another subject for punishment for more than one period. This rules out motivations such as revenge. Subject *i*'s earnings per period are now given by:

$$\left(20 - g_i + 0.4 * \sum_{k=1}^{n} g_k\right) * \frac{\max\left\{0,\ 10 - \sum_{k \neq i} P_{ki}\right\}}{10} - \sum_{k \neq i} C\left(P_{ik}\right) \qquad (2)$$

where P_{ik} is the number of points assigned by *i* to *k*, and $C(P_{ik})$ is the cost to *i* of assigning the points to *k*. The design of punishment is such that if a subject who plays the Nash equilibrium is punished, the payoff inequality is necessarily reduced between the punisher and the target. In most circumstances it guarantees a reduction of payoff inequality. [5] Losses are possible if the cost of punishing others exceeds the individual's net income but they are extremely unlikely (this situation never occurred in our experiment).

In the Equal Cost treatment, each point given by a punisher has the same monetary cost for himself and the target: the cost of being punished always equals the cost of punishing. Suppose that subject *i* assigns say 3 punishment points to subject *j*; the first-stage earnings of both subject *i* and subject *j* are reduced by 4 units, as described in Table 1. Subject *i*'s earnings per period in this treatment are given by:

$$\left(20 - g_i + 0.4 * \sum_{k=1}^{n} g_k\right) - \sum_{k \neq i} C(P_{ki}) - \sum_{k \neq i} C(P_{ik}) \qquad (3)$$

The Equal Cost treatment is tested both under a stranger matching protocol (ECS) and under a partner matching protocol (ECP).[6]

Theoretical predictions. If players are selfish, the unique subgame-perfect equilibrium of both the Unequal and Equal Cost Treatments is to contribute nothing in each period and never to punish. Punishment is not credible in either treatment. Therefore, complete free-riding is a dominant strategy in all periods and all treatments. Relaxing the selfishness assumption, consider now the predictions of Fehr and Schmidt (1999)'s inequality-aversion theory. Here individual utility depends not only on one's own payoff but also on the equality of the income distribution. Individuals are inequality averse if they incur disutility both from being worse off in material terms than others (disadvantageous inequality) and from being better off than others (advantageous inequality). Subjects are assumed to be more sensitive to disadvantageous inequality, as shown by the inequality-aversion term α_i, $\alpha_i > 0$, than advantageous inequality, given by β_i with $0 \leq \beta_i < 1$ such that $\alpha_i > \beta_i$. The utility function of player $i \in \{1,\ldots,n\}$ is:

$$U(x_i) = x_i - \alpha_i \frac{1}{n-1} \sum_{j \neq i} \max(x_j - x_i, 0) - \beta_i \frac{1}{n-1} \sum_{j \neq i} \max(x_i - x_j, 0) \qquad (4)$$

The first term of equation (4) represents the pecuniary payoff of subject *i*. The second and the third terms measure the utility loss from disadvantageous and advantageous inequality respectively.

If subjects are sufficiently averse to disadvantageous inequality, and if advantageous inequality aversion is sufficiently low relative to marginal earnings (i.e. if $1-\alpha > \beta_i$), then not contributing is

a Nash equilibrium. Now consider punishment in the Unequal Cost treatment. In the second stage, punishment strategies are credible since no enforcer gains by not punishing. Anticipating credible punishment in the second stage, no-one free rides in the first stage. Anyone deviating by free riding in stage 1 will be punished by the other group members in stage 2. Do these predictions hold when the cost of punishment is the same for the punisher and the target, as in the Equal Cost treatment? Inequality aversion predicts no punishment and no cooperation in this treatment. If a cooperator punishes a defector, whereas the other subjects play the equilibrium in the second stage, she incurs the cost of the punishment but does not change the earnings gap between herself and the target. In addition, she suffers from disadvantageous inequality relative to the cooperators who do not punish.

Procedures. The experiment was computerized using the REGATE software (Zeiliger, 2000). We ran eight sessions in total under the stranger matching protocol, split 50:50 between the Equal and Unequal Cost Treatments.[7] In total 72 subjects participated in each treatment. We also ran two sessions of the Equal Cost treatment under a partner matching protocol with 12 participants each. Six sessions were conducted in the experimental laboratory of the Groupe d'Analyse et de Theorie Economique (GATE) at the University of Lyon, and four sessions were organized in the LABoratory of EXperimental Economics at the University of Rennes, France. In total, 168 subjects were recruited from undergraduate classes in business and engineering schools in Lyon and in various departments in Rennes. We did not recruit any economics students, and none of the subjects had any experience of this particular type of experiment.

Upon arrival, the subjects drew a label from a bag, indicating the name of their computer. The instructions (see Appendix) were distributed and read aloud. The subjects then filled out a

questionnaire that allowed us to check their understanding of the rules of the game. Questions were answered in private. The program then matched subjects randomly and anonymously. Under the stranger matching protocol, groups were reshuffled after each period, whereas under the partner matching protocol, the composition of groups remained unchanged over time. During each ten-period segment, subjects did not know if the experiment would extend beyond the current segment. On average sessions lasted for 90 minutes, including reading instructions and payment. Each unit was convertible to Euro at 100 units = 2 Euros. Each participant received €16.40 on average, including a show-up fee of €3.

3. EXPERIMENTAL RESULTS

We first examine the relationship between relative income and the level of punishment. We then investigate the effects of punishment on the evolution of welfare and inequality over time.

3.1. Relative income and punishment behavior

3.1.1. Decision to punish and the intensity of punishment

In this section, we examine the distribution of punishment points by subjects to other group members and its evolution over time. Table 2 indicates the average number of punishment points distributed by treatment and their relative frequency across all periods. Figure 1 displays the evolution of the average number of punishment points over time.

[Table 2 and Figure 1: about here]

Both Table 2 and Figure 1 indicate that subjects use costly punishment in all treatments, even in the ECS treatment. We find that on average 36.4% of subjects distribute at least one punishment point in the UCS treatment, 32.2% in the ECS treatment and 37.1% in the ECP treatment. Table 2 also indicates that the distribution of points by the subjects is similar in the three treatments. Finally Figure 1 shows that in all treatments, the data exhibit the same pattern: punishment declines over time. This is stated more precisely in Result 1.

Result 1: *The impossibility to reduce payoff differences does not prevent individuals from punishing in the Equal Cost treatments.*

Support for result 1. In the stranger condition, a Mann-Whitney rank sum test accepts the null hypothesis that there is no difference in the punishment levels between the UCS and the ECS treatments (p=0.248). The same conclusion arises from the comparison between the ECS and ECP treatments (p=0.199). A more formal proof of Result 1 is given in Table 3. We have estimated various random-effects Tobit models, accounting for left and right censured observations. The dependent variable is the quantity of punishment points that player i assigns to player j in the second stage of period t. The "Equal Cost" variable is a dummy to control for possible differences between treatments in the sanctioning behavior. We also include a dummy variable to control for the matching protocol. The other independent variables include player j's contribution in period t and the average group contribution in period t. The estimations also control for negative and positive deviations from the punisher's contribution and for negative and positive deviations from the average contribution level of the group in period t. In addition, we include a time trend and dummy variable for the first period.

[Table 3 about here]

Table 3 indicates that the propensity to punish is not significantly different between the ECS and the UCS treatments and between matching protocols. The absence of any significant difference in the sanctioning behavior between the ECS and ECP treatments indicates that non strategic motives are more important than strategic motives, which is consistent with previous studies. This result also provides strong evidence that sanctions are not primarily driven by the willingness to reduce directly the current level of inequality. This result is consistent with what has been observed in prisoner's dilemma games (Falk, Fehr and Fischbacher 2005) in which people punish mostly to harm targets in retaliation for the negative emotions they have aroused. Lastly, Table 3 confirms the fact that punishment significantly declines over time.

To summarize, Result 1 concludes to the absence of differences in the punishment level across treatments. A more detailed analysis is however required to investigate some potential differences across treatments both in the dynamic and the intensity of the punishment behavior. This is stated more precisely in Results 2 and 3.

Result 2: *The punishment level in the first period is significantly higher in the partner treatment than in the stranger treatment but it is significantly lower in most of the remaining periods.*

Support for result 2. A Mann-Whitney rank sum test for the first period indicates that the level of punishment is higher in ECP than in ECS (significant at 5%). On the contrary, except for periods 16 and 17, the punishment level is significantly lower in the ECP treatment in the remaining periods (significant at 5%). Table 3 also indicates that punishment in the first period is stronger in the partner treatment than in the stranger treatments. An interpretation of this result is that punishment is more successful to lift cooperation under a partner matching protocol. For this reason, the individuals sanction more in the first period of the partner treatment, trying to

establish the credibility of sanctions from the beginning of the game. After period 1, punishment becomes credible and the threat of punishment is strong enough to maintain cooperation in the group.[8]

Turning next to the comparison between the ECS and UCS treatments, we check whether the absence of any difference between the UCS and ECS treatments pointed in Result 1 might derive from the action of two forces that work in opposite directions. On the one hand, inequality averse individuals might refrain from punishing when sanctions cannot modify the distribution of earnings. On the other hand, those who are willing to punish might distribute more points in ECS than in UCS to express their emotions, by imposing the same absolute earnings reduction than in the UCS treatment since the cost for each unit received by the target is lower in the ECS than in the UCS treatment. Indeed as shown by Falk et al. (2005) who compared a low sanction treatment with a ratio of punishment equal to one and a high sanction treatment with a ratio higher than one, "individuals increase their expenditures for punishment if the impact of given investment in punishment causes a lower payoff reduction for the punished individual." Indeed, they showed that the cooperators who punish in the low sanction treatment spent 2.5 times more on the punishment of the defectors than in the high sanction condition.[9] Result 3 confirms the existence of differences between the ECS and UCS treatments.

Result 3: *subjects are significantly less likely to punish in the Equal than in the Unequal Cost Treatment. However, conditional on the willingness to punish, subjects who punish defectors do so more intensely in the Equal than in the Unequal Cost Treatment.*

Support for Result 3: we find that, irrespective of the level of punishment, most individuals are reluctant to punish in the ECS treatment compared to the UCS treatment. Punishment is inflicted

in 12% of the cases in ECS compared to 16% in UCS . Moreover, the threshold of deviation in contribution above which participants start to punish free riders is also higher in ECS (3.70) than in UCS (1.33). This difference is statistically significant ($p<0.05$). In addition, we ran additional estimations to dissociate the decision to punish from the intensity of punishment. We consider two separable decisions using a two-step estimation procedure : first the decision to sanction someone and second, conditional on the decision to sanction, the choice of the intensity of punishment.[10] We first estimate the punishment probability using a random-effects Probit model; we then explain the number of points distributed, conditional on the decision to punish, by means of a Generalized Least Squares model corrected for a potential selection bias via the inverse of the Mills ratio (the "IMR" variable). Each regression allows us to measure the influence of the Equal Cost relative to the Unequal Cost Treatment in the stranger matching protocol. The exogenous variables in the selection equation include the positive and negative deviations between the target's contribution and both the punisher's and average group contributions[11], as well as a time trend. The GLS regression includes the same variables except for the time trend, which allows us to identify the model. The estimation results are shown in Table 4. Column (1) presents the results from the selection equation and column (2) the marginal effect of each explanatory variable. Column (3) displays the results of the GLS estimation for the whole population. The last two columns give the results of the GLS estimations for the sub-samples of observations in which the subject punishes a group member who contributes less than himself (column (4)) or more than himself (column (5)).

[Table 4: about here]

Controlling for potential selection bias and for relative contribution levels, the regressions show that, first, subjects are significantly less likely to punish in the Equal than in the Unequal Cost Treatment and, second, that those who decide to punish do so more intensely in the Equal Cost treatment. This shows that inequality aversion does play some role, since some subjects do not punish when payoff shares cannot be altered (see model (1)). However, the marginal effect of the treatment is small: the Equal Cost Treatment reduces the probability of punishment by 4% (see model (2)), which suggests a rather low number of inequality-averse subjects.

The model (3) shows that, conditional on the willingness to punish, subjects who were not inequality-averse are willing to pay more to increase the harm imposed on targets in the ECS than in the UCS treatment. This is because the monetary consequence of each punishment point on the target is lower in the ECS treatment. Indeed, on average the monetary consequence of one punishment point is 2.5 times lower in the Equal than in the Unequal Cost Treatment. Therefore, while Anderson and Putterman (2006) have shown that punishment behavior obeys the law of demand, individuals take into account not only the cost to themselves but also the cost imposed on the target. They are willing to pay a higher price in the ECS treatment to increase the monetary consequences of their sanction on the targets.[12] This result opposes to Egas and Riedl (2005) who find that there is more punishment if punishment has higher impact. In contrast, our result are consistent with Falk, Fehr and Fischbacher (2005) who observe that cooperators increase their punishment expenditures if the impact of punishment on the targets is reduced.

To sum, beyond some differences between treatments, our results show that people punish even in the Equal Cost treatment and that most of punishment behavior seems to be explained by

emotions. Next, we examine the relationship between the distribution of punishment points and the level of inequality within a period.

3.1.2. Punishment behavior and the current level of inequality

We test here the hypothesis that the observation of inter-individual differences in contributions and earnings triggers sanctions, irrespective of the ability to reduce inequality. Indeed if people allocate punishment points even when these points cannot affect the earnings gap with the punished, this does not necessarily mean that they do not care about the level of inequality. Figure 2 displays the distribution of punishment points in the second stage (on the vertical axis) as a function of the inequality of contribution and earnings between the punisher and the target at the end of the first stage (on the horizontal axis), by treatment and protocol.[13]

[Figure 2 about here]

Figure 2 clearly shows that punishers strongly react to inter-individual differences in all treatments. Indeed, in all treatments, the intensity of punishment increases in the inequality level between the punisher and the target. For example, a subject who earns between 14 and 20 units more than the punisher receives on average 2 points of punishment in the ECS and ECP treatments and 2.5 points in the UCS treatment, whereas she receives nearly no point in all treatments if she earns the same amount as the punisher. Result 4 summarizes these findings.

Result 4: *In all treatments, the intensity of punishment is strongly correlated with the level of inequality between the punisher and her target.*

Support for Result 4: Table 3 indicates that, after controlling for the deviations from the average, in all treatments subjects are highly and significantly influenced by the observation of inter-individual differences. Indeed, in almost all estimations, the coefficients associated with both positive and negative deviations from the punisher's contribution are highly significant. As in Fehr and Gaechter. (2000), player *i* sanctions player *j* more (less) the greater the negative (positive) deviation of *j*'s contribution and earnings are from *i*'s. In the ECS treatment, the coefficient associated with a negative deviation between the target and the punisher cannot reflect a willingness to reduce inequality. This result is consistent with our conjecture that income comparisons raise negative emotions that trigger sanctions.

3.2. The effects of Punishment on Welfare and Inequality

So far, we have considered the impact of relative income on individual punishment behaviour, we next examine the impact of punishment on both absolute and relative income.

3.2.1 Punishment and Welfare

In this section we investigate the social consequences of punishment on welfare over time. Our findings are summarized in results 5 and 6.

Result 5: *Punishment affects welfare in two opposite ways. On the one hand, punishment destroys resources by imposing a direct cost on both punishers and targets. On the other hand, in all treatments except ECS, this cost is progressively offset by an improvement of welfare resulting from an increase of the free riders' contributions over time.*

Support for Result 5: Table 5 provides information regarding the average payoff in each treatment.

[Table 5 about here]

The comparison of columns 1 and 3 of Table 5 indicates that the average final earnings are higher when punishment is available in both the UCS and ECP treatments. Indeed, the average final earnings with sanction in UCS (22.5) are larger than the average earnings without sanction (21.8) (significant at $p<0.1$). The respective values in the ECP treatment are 25.7 and 22.2 (significant at $p<0.05$). Average earnings are not significantly affected by punishment in the ECS treatment ($p>0.10$). Table 5 also provides information regarding the evolution of the average final earnings over time. Disaggregated data by sets of periods reveal that welfare decreases over time when no sanction can be exerted. In contrast, it increases (or remains stable in the ECS treatment) when punishment is possible. For example, in the UCS treatment, the average final payoff decreases from 22.5 to 20.8 in the periods without punishment (columns 4 and 7), whereas it increases from 20.7 to 23.1 in the periods with punishment (columns 6 and 9).

Finally, Table 5 provides evidence of the existence of two opposite effects of punishment on welfare. As mentioned in Result 5, punishment induces a direct reduction on welfare by destroying resources of both the punisher and the target. This direct cost can be easily observed in Table 5 (columns 2 and 3) by comparing final and first-stage payoffs in periods 11-20, when sanctions are available. The average cost of punishment amounts respectively to 2.9 units in the UCS treatment, 2.2 in the ECS treatment and 2.1 in the ECP treatment.[14] This result confirms the detrimental effect of punishment that has been observed in other studies (see notably Houser, Xiao, McCabe and Smith (2005)). Table 5 however indicates that this cost is decreasing over

time. Indeed, Table 5 shows that the average cost decreases from 4.8 in the first three periods to 2 in the last three periods in the UCS treatment. The corresponding values are 3.6 and 2.2 in the ECS and 3.3 and 1.5 in the ECP treatment. Further evidence of this is also provided by Figure 1.

The direct cost induced by punishment is progressively offset by a positive effect of sanctions on welfare, through an improvement of cooperation over time. The evidence of this positive effect is given in Table 5 by comparing first stage payoffs in periods 11-20 (with punishment) and payoffs without punishment over periods 1-10 and 21-30. This comparison indicates to what extent punishment induces a positive effect on welfare, by inciting individuals to contribute more. The first-stage payoffs in periods 11-20 with sanction are higher than in the pooled data from periods 1-10 and 21-30 in UCS (significant at $p<0.1$) and in ECP (significant at $p<0.05$). Finally, in the ECS treatment, punishment also improves welfare (significant at $p<0.1$). However, this increase is less important compared with the previous treatment.

Table 6 and Figures 3 also provide further evidence of the incentive effect of punishment on welfare through its positive influence on contribution. Table 6 gives the average individual contribution to the public good in the various treatments in each set of periods. Figure 3 displays the evolution of the average contribution by period for each treatment.

[Table 6 and Figures 3: about here]

In all treatments, the data exhibit the same pattern: in periods 1-10, subjects initially contribute more than the Nash equilibrium level, but progressively reduce their contributions; in periods 11-20 the introduction of the punishment opportunity entails an increase in average contributions; last, in periods 21-30, average contributions drop off sharply when punishment opportunities are withdrawn. This is stated more precisely in Result 6.

Result 6: *In all treatments, the opportunity to punish causes an increase in average contributions. The positive effect of punishment on contribution is lower in the ECS. On the contrary, this effect is higher under a partner matching protocol (ECP treatment), confirming that punishment is more successful to lift cooperation under a partner matching protocol.*

Support for Result 6: In the ECP treatment, Table 6 shows that the average contribution is 4 when no punishment is available and rises to 13.6 in the periods with punishment. A Wilcoxon matched pair test rejects the null hypothesis that contributions are identical between periods 11-20 and the pooled data from periods 1-10 and 21-30, (p=0.03).[15] In the UCS treatment, the average contribution is 3.65 in the periods without punishment and rises to 9.59 in the periods with punishment. The Wilcoxon matched pair test also rejects the null hypothesis (p=0.07). In the ECS treatment, the average contribution is 4.06 in the periods without punishment and 7.73 in the periods with punishment. The same test also reveals a significant difference between periods 11-20 and the pooled data from periods 1-10 and 21-30 (p<0.1). Nevertheless, the increase in contribution is lower in ECS compared to the previous treatments.

Table 7 provides a more formal evidence of these results. It reports the estimates of random-effects GLS models analyzing the determinants of the change in contribution between periods t and $t+1$. These are estimated separately for those who contributed less than average and more than average (on the left and right sides of the table respectively). Columns 1 and 4 correspond to the pooled data from the UCS and ECS treatments; columns 2 and 5 correspond to the pooled data from the ECS and ECP treatment; columns 3 and 6 pool all data together. The explanatory variables include the punishment received in period t and the deviation from the average

contribution of other group members.[16] We also interact the punishment points received with the treatment and with the protocol. These interactions show whether punishment is more or less effective under the Equal Cost treatment, and under the partner matching protocol.

[Table 7 about here]

Table 7 provides interesting comparative results on the effectiveness of punishment. Indeed, it indicates that the effect of sanctions on the target strongly differs between treatments. The regressions in Table 7 show that subjects react more to punishment under a partner matching protocol, underlining the importance of strategic motives for contributing in repeated interactions. Moreover, the positive effect of punishment is significantly higher in the UCS than in the ECS treatment. This is due to the fact that, by design, one punishment point costs on average less to the target in the ECS than in the UCS treatment.[17] As a consequence, the incentive effect of punishment on further contribution is lower under the ECS treatment. This explains that the impact of punishment on the evolution of welfare over time is lower in ECS than in UCS.

3.2.2 Punishment and inequality

How do previous results translate into the dynamics of inequality both at the individual and aggregate levels? Our findings are summarized in Result 7.

Result 7: *In all treatments, sanctions impact inequality over time in two ways. On the one hand, sanctions reduce inequality by inciting free riders to contribute more. On the other hand, sanctions impact inequality by the cost of punishment. At the individual level, this cost reduces (or at least remains unchanged in the Equal Cost treatments) inequality between the punisher*

and the target whereas it generates higher payoff dispersion in all treatments between those who incurred those costs either as punishers or as targets at the aggregate level.

Support for Result 7.

Table 5 provides interesting information about inequality of income distribution both at individual and aggregate levels. As it was the case for welfare, Table 5 reflects the existence of two opposite effects of punishment on inequality of the distribution of payoffs. On the one hand, punishment impacts the dispersion of earnings by the cost of punishment. At the individual level, Table 5 shows that the inter-individual inequality between the punisher and the target decreases between first stage and final stage with sanction in the UCS treatment whereas it remains unchanged in the other treatments. In contrast, at the aggregate level, Table 5 indicates that sanctions increase inequality of the distribution of payoff between those who punish or/and are punished and those who neither punish nor are punished. We consider here two different measures of inequality at the aggregate level: (a) the Gini coefficient as a measure of relative inequality, (b) the standard deviation as a measure of absolute inequality .[18] Table 5 confirms this by showing that the comparison of the Gini coefficients between the final payoffs and the first-stage payoffs when sanctions are available indicates that the costs associated with sanctions tend to increase the dispersion of earnings. The Gini coefficient for final payoff with punishment is 0.11 in UCS and ECS and 0.10 in ECP. A Wilcoxon matched pair test rejects, for all treatments, the null hypothesis that the Gini coefficient is similar between the first-stage and final payoffs (significant at $p<0.1$). Similar results are found using standard deviation as an absolute inequality measure. A Wilcoxon matched pair test also rejects the null hypothesis that the standard deviation coefficients are similar between the first-stage and final payoffs, (significant at $p<0.1$).

On the other hand, punishment tends to reduce the dispersion of earnings by inciting all individuals to converge toward a similar strategy of cooperation. Table 5 indicates that the Gini coefficients associated with the average payoffs without sanction are higher (0.10) than the similar coefficients for the first-stage payoffs when punishment is available (0.09), in the UCS and ECS treatments (significant at $p<0.1$). A Wilcoxon matched pair test also rejects the null hypothesis that the Gini coefficient is similar between the first-stage payoff with punishment and the payoff without punishment in the ECP treatment (significant at $p<0.1$). Similar results are found using standard deviation. If we compare the evolution of inter-individual inequality with and without sanctions (columns 1 and 2), we observe that inequality at the individual level also decreases when sanctions are available in all treatments. Finally, Table 5 shows that inequality declines over time within each set of periods in each treatment. For example, in the UCS treatment, the Gini coefficients decline from 0.12 to 0.06 without punishment, and from 0.13 to 0.10 with punishment. Standard deviation coefficients also decline over time in all treatments. The values for ECS are 0.11 to 0.08 without punishment and 0.13 to 0.11 with punishment, respectively. These results support an important finding: when punishment is not possible, free riding leads to more equal payoffs but with falling earnings; on the contrary, punishment produces falling inequality along with an improved welfare.

4. DISCUSSION AND CONCLUSION

In this paper we have investigated the complex relationship between welfare and punishment decision. In particular, we have explored to what extent sanctions are driven by a willingness to reduce payoff differences. Indeed, the experimental literature has established the

importance of inequality aversion in shaping social preferences and behavior. In their inequality aversion theory, Fehr and Schmidt (1999) suggest that altruistic punishment can be motivated by the willingness to reduce earnings differentials. To assess the importance of this motivation, we have considered in this paper a number of treatments of a public good game in which we introduce punishment opportunities with different consequences on the distribution of payoffs. In particular, we have compared an environment where costly punishment cannot affect the distribution of payoffs (Equal Cost treatment) with a standard two-stage public good game in which the monetary consequences of sanctions are larger for the target than for the punisher (Unequal Cost treatment). In the first environment, an individual only motivated by the willingness to reduce the current level of inequality should not punish. We have tested this prediction with both a stranger and a partner matching protocols. In addition, we have analyzed to what extent punishment affects welfare and inequality over time. Indeed, even if sanctions do not aim at immediately reducing payoff differences, they may affect both welfare and the distribution of earnings over time. We have examined two opposite effects of punishment both on welfare and inequality. On the one hand, punishment negatively affects welfare and increases the dispersion of earnings by imposing direct costs on both the punisher and her target. On the other hand, sanctions increase welfare and reduce inequality at the aggregate level, by inciting individuals to contribute more over time.

We have four key findings. First, individuals punish even when punishment cannot affect immediately the distribution of payoffs. Therefore, consistent with previous studies, our results indicate that punishers are not primarily motivated by the willingness to reduce the current level of inequality between themselves and the targets. Second, we argue that, in all treatments, inter-individual comparisons play a decisive role in the decision to punish free riders. Indeed, we

show that, in all treatments, the intensity of punishment is strongly correlated with the extent of the difference in contributions and earnings between the punisher and his target. This result indicates that, irrespective of the willingness to *directly* reduce payoff differences, individuals may be willing to punish those whose decisions give rise to payoff differences, and that such payoff differences arouse emotions that trigger punishment.

The last two results are related to the dynamic impact of sanctions on both welfare and inequality. We find that punishment affects both welfare and inequality in two opposite ways. On the one hand, it imposes a direct cost on both the punisher and the target, which exerts a negative influence on welfare and tends to increase the dispersion of earnings at the aggregate level. On the other hand, punishment exerts a positive influence on welfare and reduces the dispersion, by inciting individuals to converge toward cooperation. The aggregate effect of punishment on the evolution of welfare reveals to be positive both in the Unequal Cost treatment and in the Equal Cost treatment with a partner matching protocol. The disciplinary effect of punishment is stronger under a partner protocol where the threat of future sanctions is more credible. Finally, welfare is constant in the Equal Cost treatment with a stranger matching protocol. Indeed the incentive effect of punishment is less strong in this treatment due to the lower consequences of one punishment point on the target's payoff. In contrast, in all treatments, welfare declines sharply when punishment is not available. Finally, when we consider both the evolution of welfare and the evolution of inequality at the aggregate level, we show that in the absence of punishment opportunities, free riding develops and progressively gives rise to reduced inequality coupled with low earnings. In contrast, when a right to punish is implemented, the disciplinary effect of punishment brings about falling inequality over time associated with higher earnings.

These results open new perspectives for further research works, related to the individual determinants of punishment behavior and to the impact of punishment on welfare. In particular, it would be interesting to analyze how various punishment institutions would allow for an improvement of welfare by reducing the detrimental effect of sanctions.

REFERENCES

Anderson, Christopher M. and Putterman, Louis, (2006). "Do Non-Strategic Sanctions Obey the Law of Demand? The Demand for Punishment in the Voluntary Contribution Mechanism." *Games and Economic Behavior*, *54*(1), 1-24.

Besançon, M.L. (2005). Relative resources: inequality in ethnic wars, revolutions and genocids". *Journal of Peace Research*, 42(4), 393-415.

Bochet, Olivier; Page, Talbot and Putterman, Louis, (Forthcoming). "Communication and Punishment in Voluntary Contribution Experiments." *Journal of Economic Behavior and Organization.*

Bolton, Gary E. and Ockenfels, Axel, (2000). "ERC: A Theory of Equity, Reciprocity, and Competition." *American Economic Review*, *90*(1), 166-93.

Bowles, Samuel and Herbert Gintis (2000). "The Evolution of Strong Reciprocity", Working paper, Santa Fe Institute.

Brown, Gordon D.A.; Gardner, Jonathan; Oswald, Andrew J. and Qian, Jing, (2005). "Does Wage Rank Affect Employees' Wellbeing?" *IZA Discussion Paper*, 1505. Bonn.

Carpenter, J.P., (2006). "Punishing Free-Riders: How Group Size Affects Mutual Monitoring and the Provision of Public Goods." *Games and Economic Behavior*, (Forthcoming).

Carpenter, J.P., (2007). " The demand for punishment" *Journal of Economic Behavior & Organization*. 62 522–542

Carpenter, J.P.; Matthews, P. and Ong'ong'a, O., (2004). "Why Punish? Social Reciprocity and the Enforcement of Pro-Social Norms." *Journal of Evolutionnary Economics*, *14*.

Charness Gary and Matthew Rabin (2002). "Understanding Social Preferences with Simple Tests" *Quarterly Journal of Economics*, 117(3), 817-869.

Clark, Andrew E. and Oswald, Andrew J., (1996). "Satisfaction and Comparison Income." *Journal of Public Economics*, 61, pp. 359-81.

Egas, Martin, and Riedl, Arno, (2005). "The Economics of Altruistic Punishment and the Demise of Cooperation." *IZA Discussion Paper* No.1646.

Falk, Armin; Fehr, Ernst and Fischbacher, Urs, (2002). Testing Theories of Fairness - Intentions Matter, University of Zürich, Working Paper No. 63.

Falk, Armin; Fehr, Ernst and Fischbacher, Urs, (2005). "Driving Forces Behind Informal Sanctions." *Econometrica*, 73 (6), 2017-2030.

Falk, Armin and Fischbacher, Urs (2006), "A Theory of Reciprocity", *Games and Economic Behavior*, 2006, 54 (2), 293-315.

Fehr, Ernst and Gächter, Simon, (2002). "Altruistic Punishment in Humans." *Nature*, *415*(10), 137-40.

____, (2000). "Cooperation and Punishment in Public Goods Experiments." *American Economic Review*, *90*(4), 980-94.

Fehr, Ernst and Rockenbach, Bettina, (2003). "Detrimental Effects of Sanctions on Human Altruism." *Nature*, *422*, 137-40.

Fehr, Ernst and Schmidt, Klaus M., (1999). "A Theory of Fairness, Competition and Cooperation." *Quarterly Journal of Economics*, *114*, 817-68.

Ferrer-i-Carbonell, A., (2005). "Income and Well-Being: An Empirical Analysis of the Comparison Income Effect." *Journal of Public Economics*, 89, pp. 997-1019.

Hopfensitz, Astrid and Reuben, Ernesto, (2005). "The Importance of Emotions for the Effectiveness of Social Punishment." *Tinbergen Institute Discussion Paper*, TI 2005 - 075/1. Amsterdam.

Houser, Daniel; Xiao, Erte; McCabe, Kevin and Smith, Vernon, (2005). "When Punishment Fails: Research on Sanctions, Intentions and Non-Cooperation." *George Mason University Working Paper*.

Lowery Daniel and Sigelman Lee, (1981). "Understanding the Tax revolt : eight Explanations." The Americal Political Science Review, *75* (4), 963-974.

MacCulloch, R., and Pezzini, S. (2004). "The Role of Freedom, Growth, and Religion in the Taste for Revolution". Working Paper, London School of Economics.

____, (2007). "Money, religion and revolution". Economics of Governance, 8(1), 1-16.

Masclet, David; Noussair, Charles; Tucker, Steve and Villeval, Marie-Claire, (2003). "Monetary and Non-Monetary Punishment in the Voluntary Contributions Mechanism." *American Economic Review*, *93*(1), 366-80.

Neumark, David and Postlewaite, Andrew, (1998). "Relative Income Concerns and the Rise in Married Women's Employment." *Journal of Public Economics*, *70*, pp. 157-83.

Nikoforakis, Nikos, and Normann, Hans-Theo, (2005). "A Comparative Statics Analysis of Punishment in Public-Good Experiment.", Royal Holloway, University of London, mimeo.

Nikoforakis, Nikos, Normann, Hans-Theo, Wallace, Brian (2005). "Asymmetric Punishments in Public-Good Experiments.", Royal Holloway, University of London, mimeo.

Price, Michael E.; Cosmides, Leda and Tooby, John, (2002). "Punitive Sentiment as an Anti-Free Rider Psychological Device." *Evolution and Human Behavior*, *23*, 203-31.

Quervain, D.J.F.; Fischbacher, Urs; Treyer, V.; Schellhammer, M.; Schnyder, U.; Buck, A. and Fehr, E., (2004). "The Neural Basis of Altruistic Punishment." *Science*, *305*, 1254-58.

Rabin, Matthew, (1993). "Incorporating Fairness into Game Theory and Economics." *American Economic Review*, *83*, 1281-1302.

Wright, S.C., Donald M. Taylor, D.M., and Moghaddam, F.M. (1990). "The relationship of perceptions and emotions to behavior in the face of collective inequality" *Social Justice Research*, 4(3), 229-250.

Zeiliger, Romain, (2000). *A Presentation of Regate, Internet Based Software for Experimental Economics.* http://www.gate.cnrs.fr/~zeiliger/regate/RegateIntro.ppt., GATE.

Table 1. Levels of punishment and associated costs

Unequal Cost treatment

Punishment Points	0	1	2	3	4	5	6	7	8	9	10
Cost to the punisher in units	0	1	2	4	6	9	12	16	20	25	30
Cost to the target in % of the target's earnings from the 1st stage	0	10	20	30	40	50	60	70	80	90	100

Equal Cost treatment

Punishment Points	0	1	2	3	4	5	6	7	8	9	10
Cost to the punisher in units	0	1	2	4	6	9	12	16	20	25	30
Cost to the target in units	0	1	2	4	6	9	12	16	20	25	30

Table 2. Distribution of punishment points by treatment

Treatment	Average number of points by subject & by period	Relative frequency of points (in percentage)				
		0 point	1-2 points	3-4 points	5 points and more	Total
UCS	0.7 (1.2)	63.6	26.0	9.0	1.4	100
ECS	0.9 (1.7)	67.8	20.0	7.2	5.0	100
ECP	0.9 (1.6)	62.9	26.3	7.9	2.9	100

Note: Standard deviations in parentheses

Table 3. Determinants of sanctioning behavior

	Random-effects Tobit models			
Variables	Pooled data (1)	UCS (2)	ECS (3)	ECP (4)
Equal cost punishment	-0.145 (0.281)			
Partner matching protocol	0.472 (0.366)			
Negative dev. From punisher's contribution	0.509*** (0.018)	0.412*** (0.022)	0.764*** (0.049)	0.418*** (0.043)
Positive dev. From punisher's contribution	-0.092*** (0.025)	-0.032 (0.027)	-0.110* (0.066)	-1.140*** (0.298)
Negative deviation from the average	0.043 (0.029)	0.038 (0.035)	-0.017 (0.073)	-0.250* (0.146)
Positive deviation from the average	-0.321*** (0.023)	-0.216*** (0.028)	-0.427*** (0.053)	-0.489*** (0.090)
Average group contribution	-0.046*** (0.013)	-0.010 (0.021)	-0.012 (0.031)	-0.063 (0.044)
Time trend	-0.144*** (0.015)	-0.099*** (0.021)	-0.099** (0.043)	-0.085* (0.05)
Period 1		0.189 (0.173)	0.180 (0.357)	1.512*** (0.484)
Constant	-0.531 (0.354)	-0.812* (0.474)	-2.626*** (0.822)	0.079 (1.22)
Observations	5760	2160	2160	720
Left censored obs.	4845	1797	1875	613
Log Likelihood	-2864.40	61029.49	61030.28	6336.50
Wald Chi2	1116.21	504.12	328.19	134.94
Prob>Chi2	0.000	0.000	0.000	0.000

Note: Standard errors in parentheses; *significant at 10%; ** significant at 5%; *** significant at 1%

Table 4. Punishment (probability and intensity) in the Stranger Matching protocol

Variables	Random effects Probit model		GLS models		
	(1)	(2)	(3)	(4)	(5)
Equal Cost Treatment	-.416*** (.140)	-.036***	.488*** (.121)	.493*** (.134)	-.099 (.199)
Negative difference from the average	.022 (.023)	.002	-.021 (.187)	-.023 (.045)	.007 (.038)
Positive difference from the average	-.227*** (.020)	-.019***	-.037* (.023)	-.040* (.023)	.843*** (.154)
Average group contribution	-.032** (.013)	-.003**	-.015 (.013)	-.029** (.014)	-.003 (.034)
Negative difference from the punisher	.361*** (.016)	.031***	.112*** (.029)	.101*** (.030)	
Positive difference from the punisher	-.042** (.018)	-.004**	.085*** (.024)		.046 (.032)
Period	-.114*** (.013)	-.010***			
Constant	.180 (.276)		1.224*** (.310)	1.519*** (.327)	.205 (.897)
ρ	.496 (.040)				
IMR			-.210* (.127)	-.308** (.132)	.322 (.415)
Nb observations	4320		668	597	59
Log Likelihood	-1119.615				
R^2			.258	.261	.551
Wald χ^2	711.29		238.77	242.54	64.17
p> χ^2	.0000		.0000	.0000	.0000

Note: standard errors in parentheses. *** statistically significant at the .01 level; ** at the .05 level; * at the .10 level. Each individual appears 30 times (1 observation for each pair within the group over 10 periods)

Table 5. Average payoffs and inequality by treatment

	All periods			First periods			Final periods		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Groups of periods	Without sanction	With sanction 1-stage payoff	With sanction Final Payoff	Without sanction 1-3 and 21-23	With sanction 11-13 1-stage payoff	With sanction 11-13 final payoff	Without sanction 8-10 and 28-30	With sanction 18-20 1-stage payoff	With sanctio 18-20 final payoff
UCS treatment									
Mean Payoff	21.8	25.4	22.5	22.5	25.5	20.7	20.8	25.1	23.1
Interpers.Inequ. (abs. value)#	*5.38*	*4.95*	*4.89*	*8.68*	*5.97*	*5.8*	*4.29*	*4.20*	*4.10*
Aggregate level Inequality									
Sdt. Dev.	(4.16)	(4.13)	(4.52)	(4.76)	(4.63)	(4.90)	(3.31)	(4.20)	(4.30)
Gini coef	[0.10]	[0.09]	[0.10]	[0.12]	[0.10]	[0.13]	[0.06]	[0.09]	[0.10]
ECS treatment									
Mean Payoff	22.1	24.2	22.0	23	25.5	21.9	21.2	23.1	21
Interpers.Inequ. (abs. value)#	*5.54*	*5*	*5*	*6.57*	*5.76*	*5.76*	*5.08*	*4.63*	*4.63*
Aggregate level Inequality									
Sdt. Dev.	(4.38)	(4.10)	(5.08)	(4.96)	(4.46)	(5.90)	(3.46)	(3.91)	(4.97)
Gini coef	[0.10]	[0.09]	[0.10]	[0.11]	[0.09]	[0.13]	[0.08]	[0.08]	[0.11]
ECP treatment									
Mean Payoff	22.2	27.8	25.7	22.6	27.4	24.1	20.7	27.2	25.7
Interpers.Inequ. (abs. value)#	*5.57*	*5.25*	*5.25*	*6.93*	*6*	*6*	*5.51*	*4.64*	*4.64*
Aggregate level Inequality									
Sdt. Dev.	(4.69)	(4.09)	(5.62)	(5.11)	(5.03)	(7.06)	(4.52)	(5.30)	(5.38)
Gini coef	[0.11]	[0.08]	[0.10]	[0.11]	[0.09]	[0.11]	[0.06]	[0.08]	[0.11]

Interpretation : Interpersonal inequality is calculated as the difference of payoff between player *i* and player *j*. Under t assumption of self-centered inequality, subject *i*'s final earnings in a period of the Unequal Cost Treatment is given $\left(20-g_i+0.4*\sum_{k=1}^{n}g_k\right)-C\left(P_{ij}\right)$. Subject j' s final earning is given by $\left(20-g_j+0.4*\sum_{k=1}^{n}g_k\right)*\frac{\max\left\{0,\ 10-P_{ij}\right\}}{10}$. Subject i's fir earnings in the Equal Cost treatments is given by : $\left(20-g_i+0.4*\sum_{k=1}^{n}g_k\right)-C\left(P_{ij}\right)$. Subject *j*'s final earnings in the Equal Cc Treatments is given by $\left(20-g_j+0.4*\sum_{k=1}^{n}g_k\right)-C\left(P_{ij}\right)$.

Table 6. Average individual contributions across treatments

Treatment	Periods 1-10 (without punishment)	Periods 11-20 (with punishment)	Periods 21-30 (without punishment)	Periods 1-10/21-30 (Pooled data)
UCS	5.0	9.6	2.3	3.7
	(5.4)	(5.4)	(3.7)	(4.8)
ECS	6.3	7.7	1.8	4.1
	(5.9)	(5.2)	(3.3)	(5.3)
ECP	4.0	13.6	4.8	4.4
	(5.3)	(6.2)	(5.8)	(5.5)

Note: Standard deviations in parentheses

Table 7. Determinants of the change in contributions between *t* and *t+1*

Dependent variable: Change in contribution between *t* and *t+1*	Targets who contributed less than the average			Targets who contributed more than the average		
	Stranger (1)	Equal (2)	Pooled data (3)	Stranger (4)	Equal (5)	Pooled data (6)
Points received in period *t*	0.894*** (0.126)	0.506*** (0.104)	0.833*** (0.125)	0.653* (0.388)	0.220 (0.523)	0.631 (0.414)
Points received * ECS	-0.0412*** (0.111)		-0.376*** (0.116)	-0.448 (0.597)		-0.441 (0.639)
Points received * partner matching		0.251* (0.146)	0.238* (0.130)		1.660 (1.823)	1.630 (1.742)
Deviation from the average	-0.314*** (0.060)	-0.303*** (0.077)	-0.349*** (0.055)	-0.770*** (0.054)	-0.690*** (0.077)	-0.713*** (0.054)
Constant	-0.522** (0.215)	-0.398 (0.317)	-0.492** (0.211)	0.277 (0.288)	-0.214 (0.376)	-0.002 (0.280)
Observations	616	380	692	634	434	744
R^2	0.308	0.276	0.322	0.207	0.150	0.166
Wald χ^2	287.68	157.63	352.27	203.82	82.49	174.69
p> χ^2	0.000	0.000	0.000	0.000	0.000	0.000

Note: Standard errors in parentheses. *** statistically significant at the .01 level; ** at the .05 level: * at the .10 level.

Figure 1. Average number of punishment points distributed over time

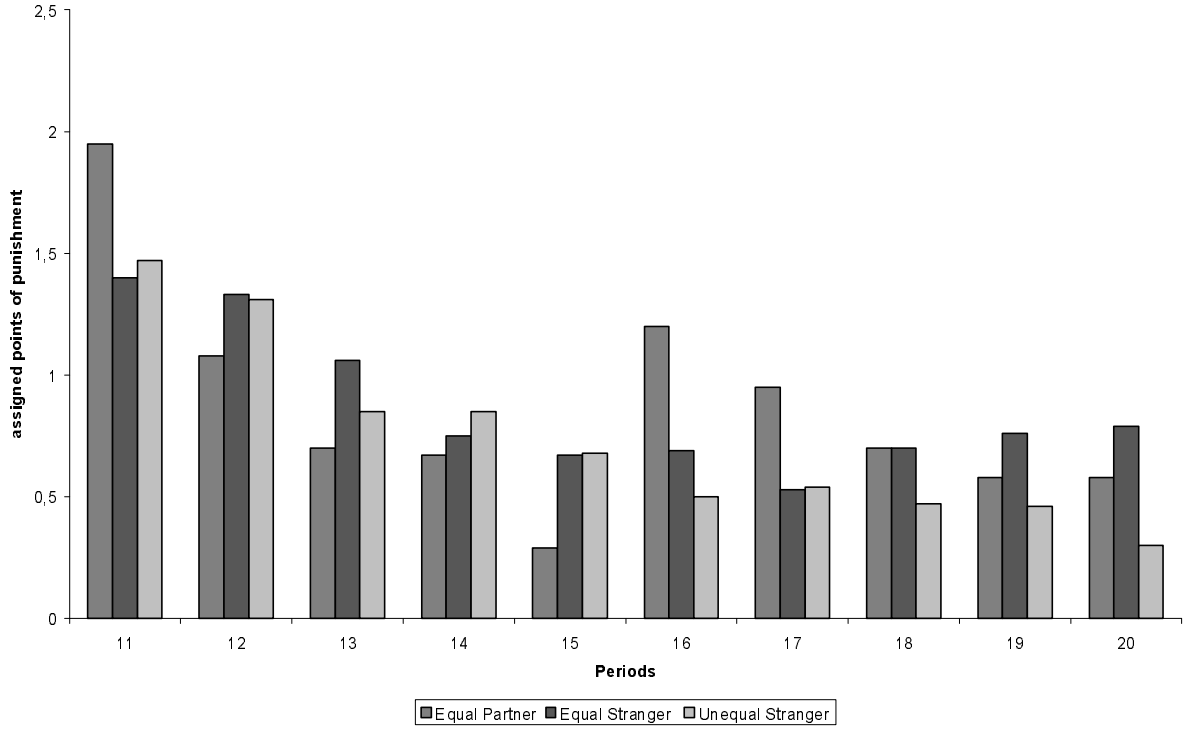
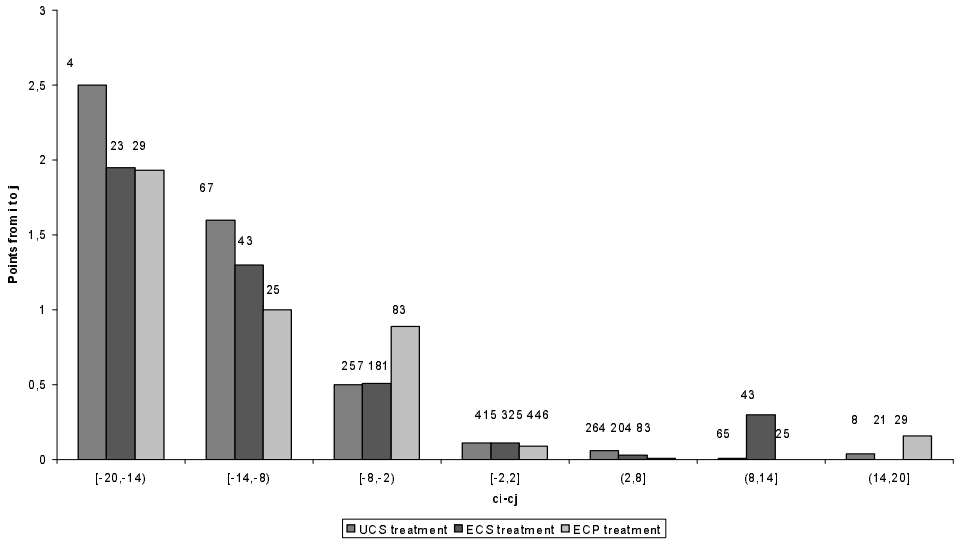
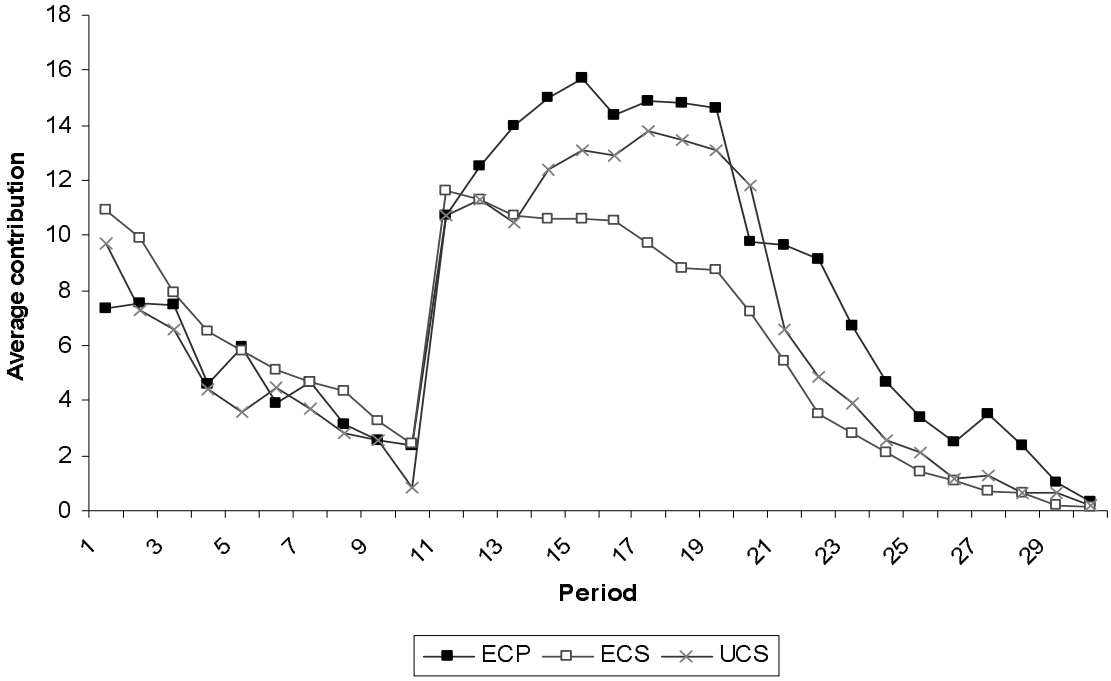


Figure 2. Punishment as a function of contribution and earning inequality between the punisher and the target



Note: The percentages at the top of the bars represent the relative frequency of each category of deviation in the various treatments.

Figure 3. Individual contribution level per treatment



NOTES

[1] Some models sought to isolate distributional concerns from intentions (Charness and Rabin, 2002; Falk, Fehr and Fischbacher, 2002). Charness and Rabin (2002) show that subjects are more concerned with intentions and the willingness to increase social welfare than with reducing differences in payoffs. Using an experimental design that allows to test these two theories, Falk et al. (2000) observed that intentions play an important role and that equity models cannot fully explain all reciprocal behaviors. The fitness differential theory in evolutionary psychology also posits that punishment aims at reducing the payoff advantage of free-riders (Price, Cosmides and Tooby 2002). In evolutionary economics see also Bowles and Gintis (1999).

[2] In the first stage of the game designed by Fehr and Gächter (2000), subjects contribute to a public good; in the second stage, after being informed about their individual contributions, the subjects can impose costly punishment on their team members. Contrary to the unique sub-game perfect Nash equilibrium of this game, subjects do punish their teammates whose level of contribution is lower than the average. The targets increase their contributions in reaction to punishment and the groups converge to the optimum of full cooperation.

[3] In the Unequal Cost treatment, this ratio is theoretically greater than one in most circumstances In our experiment, the cost to the target is higher than the cost to the punisher in all the cases where a subject punishes another subject in the Unequal Cost treatment. On average, the ratio was 2.74.

[4] In contrast with Falk, Fehr and Fischbacher (2005), we consider here another type of cooperation game, a four-person public-good experiment, in which the set of actions is larger (including a larger set of contribution levels but also of punishment levels). There are several differences between our design and the previous study of Falk et al. (2006): first we used a public good experiment instead of a prisoner dilemma game. Our game provides a continuum of strategies, which allows us to measure to what extent deviation from contribution are punished. Second our experiment allows us to study the dynamic of the game since the game is repeated (instead of a one shot game in Falk et al.). Third, the decision of sanction consists of a continuous schedule of punishment (instead of a binary decision), which allows us to measure the expenditure of punishment for each deviation.

[5] Our Unequal Cost treatment is built on Fehr and Gaechter (2000) who used a convex cost of punishment and a percentage reduction of the income of the punished player. We acknowledge that one property of the study is that by construction the punishment is stated in percentages terms in the UCS treatment and in absolute numbers in the ECS and ECP treatments. Such difference may generate a potential framing effect. Alternatively, we could have imposed a fixed cost ratio between the punisher and the target higher than 1. However note that the main purpose of our study was not to compare different levels of ratio of punishment but rather compare a situation where the ratio equals one (ECS) with a situation where the ratio of punishment is higher than one, irrespective of the level of this ratio. For more detail on the literature on the demand for punishment, see Egas and Riedl (2005), Nikoforakis, and Normann, (2005), Carpenter (2006), Anderson and Putterman (2006). Fehr and Schmidt (1999) note that average punishment effectiveness in Fehr and Gaechter (2000) is 3. In our experiment the ratio of punishment was always higher than one in all situations. On average, this ratio was 2.74. Finally the main purpose of our study was to investigate the relationship between welfare and punishment in particular when the ratio of punishment equals one. In this sense, the comparison between the ECS and UCS treatments is more illustrative than informative.

[6] By construction, the impact of receiving points, for a given period, is on average lower in the Equal Cost treatment. Consider the following examples. Suppose that player *j*'s first-stage payoff is 20 ECU and that player i assigns her 2 points. Then, player *j*'s first-stage payoff is reduced by 4 ECU in the Unequal Cost treatment while it is reduced by 2 ECU only in the Equal Cost treatment. Consider now the case where player *j*'s first-stage payoff is 32 units. For the same amount of received points, player *j*'s first-stage payoff is reduced by 6.4 units in the Unequal Cost treatment and by 2 units in the Equal Cost treatment. Finally, suppose that player's *j* first-stage payoff is 44 units. Then, player *j*'s first-stage payoff is reduced by 8.8 ECU in the Unequal Cost treatment whereas it is only reduced by 2 units in the Equal Cost treatment.

[7] Due to the number of repetitions, the stranger matching protocol cannot avoid meeting a same subject several times during a session. If this does not allow ruling out strategic considerations completely, we observe that our results are quite similar to those obtained by Fehr and Gächter (2000) under a perfect stranger matching protocol.

[8] We also ran additional estimations with a dummy variable for each period (available upon request) showing that the level punishment in later periods is weaker in the partner treatment.

[9] The literature on punishment has investigated how punishment differs both in the cost and impact of punishment. While the conclusions that emerge from the majority of these researches indicate a strong inverse relationship

between the cost of punishment and the demand for punishment there is however no clear consensus on the effect of the impact of punishment. Falk et al (2006) observe that individuals increase their expenditures for punishment if the impact of given investment in punishment causes a lower payoff reduction for the punished individual. On the contrary, Egas and Riedl (2005) observe that there is more punishment if it is cheaper but also if its impact is higher.

[10] In the context of a public good game with sanction, Carpenter and Peter Hans (2007) show that that the decision to punish should be modelled separately from the decision of how much to punish. Indeed the authors suspect that "for most individuals, the decisions whether or not to punish and how much to punish were not just two sides of the same coin".

[11] The "negative deviation from punisher" variable is the absolute value of the difference between subject *j* 's contribution and the contribution of subject *i*. This variable is set equal to zero if the deviation is positive. The other deviation variables are constructed analogously.

[12] Interestingly, this is only true for the subjects who punish those who contribute less than themselves (model (4)). In contrast, the subjects who punish those who cooperate more and are not inequality averse, by definition, are not willing to pay more to punish in ECS than in UCS (model (5)). An interpretation is that people feel negative emotions when measuring their distance to others but the intensity of these emotions is lower than that of the cooperators *vis-à-vis* free-riders.

[12] In standard linear public good games, the difference in contributions between two players is similar to the difference between their first-stage payoffs.

[13] No distinction is made here between the cost incurred by the punished and those incurred by the target.

[14] In all non-parametric statistical tests reported in this paper, the unit of observation is the group in the data collected under the partner matching protocol (N=6) and the session for the data collected under the stranger matching protocol (N=8).

[15] Here we consider the deviation from the average since a punished subject is only aware of the total number of points she received from the other group members and does not know who punished him.

[16]The average cost to the target per punishment point is 2.44 units in UCS, 1.13 units in ECS and 1.10 units in ECP.

[18] Indices of inequality are generally used to evaluate and compare different income distributions. Well-known examples are the Gini-coefficient, the standard deviation, the coefficient of variation, and Theil's measure. In the literature, two type of measures are considered : measures of absolute and relative inequality. Measures of relative inequality (or "rightist" measures) are not changed by equiproportional variations of all incomes; whereas, measures of absolute inequality or "leftist" measures are invariant with respect to equal additions. The most prominently used relative measure of statistical dispersion is the Gini coefficient. This inequality measure is a relative inequality measure since, if all incomes grow at the same rate, the relative measure of inequality remains constant (indicating a constant degree of inequality). In order to investigate whether our results hold for inequality measures other than the Gini coefficient, we also use the standard deviation which is an absolute measure of inequality.